国际哲学与科学交叉学科研究进展评述

刘 闯[1*] 朱科夫[2]

1 复旦大学 哲学学院 上海 200433

2 中国科学院大学 人文学院 北京 100049

摘要 文章介绍与评论了在高能物理学、智能科学、生物科学和社会科学等学科的前沿领域中，哲学与科学跨学科研究的近况。在简要说明了这些学科的哪些前沿问题与哲学发生了相互碰撞的同时，也介绍了近年来这些方面在国际上出现的交叉研究机构与人物。

关键词 交叉学科，黑洞物理学，早期宇宙物理学，量子引力理论，预测处理模型，自由能定律，演化博弈论

DOI 10.16418/j.issn.1000-3045. 20210115001

按照今天人们一般的理解，哲学与科学是两个截然不同的概念，这两大学科也应该是两个少有相互往来和交流的学术共同体。这种状况，至少在西方文明中，其实直到现代才出现。远的不说，即使到了文艺复兴之后，哲学与科学（甚至实验科学）都是相互渗透、同步发展的。两者之间的这种紧密关系不但体现在同一个人的思想和研究中，如笛卡尔和莱布尼兹；而且亦体现在哲学家与科学家的交流与友谊上，如在洛克与牛顿之间。这种相互关系产生出了近代科学与哲学的辉煌成就[1,2]。这一传统在西方一直以某种程度延续着，以至于到了20世纪初，马赫的哲学思想还对爱因斯坦发现相对论有过直接的影响[3]。

现代西方哲学与科学的分离是诸多因素的结果，其中较为重要的因素包括英国的工业革命与扩张殖民，西方哲学的中心由英国和法国向科学技术相对落后的德国转移，以及浪漫主义思潮对德国哲学的推进等[4]。哲学中心南移至德语民族的势头如此强烈，以至于到了19世纪末、20世纪初，在剑桥大学和牛津大学的哲学系中最有权威的哲学家均为黑格尔哲学的忠实解读者。

科学哲学可以说是离科学最近的哲学，尤其是其中包括了数学哲学、物理学哲学等与具体的科学分支

*通讯作者

修改稿收到日期：2021年1月15日

紧密结合的方向。可是科学哲学在20世纪初的兴起，最开始却是为了摆脱当时黑格尔哲学的统治，以科学的精神改造哲学。在其后1个多世纪里，从逻辑、数学和自然科学中借鉴的问题与方法彻底改变了西方哲学的面貌[5]。然而，大多数做科学研究的学者，对哲学既不了解也不关心；科学哲学逐渐演变成了一个独立于科学的人文学科——或是注重自身产生的哲学问题，或是研究科学实践的社会文化层面，“内卷”的现象日趋加剧。

在二者分离之前，科学与哲学曾被统称为“自然哲学”。牛顿力学及其时空物质框架与爱因斯坦相对论及其时空物质框架，无疑是自然哲学的典范。它们出现的时代都是人类对自然奥秘的探索遇到根本性问题和困难的时代。16、17世纪诸多由实验方法发现的自然现象，已经无法在基督教与自然机械论的二元解释框架中找到归属；同样地，19世纪末以牛顿力学绝对时空观为背景的电动力学在理论与实验之间的相悖，也是当时物理学出现危机的原因。今天，科学许多领域的前沿研究所遇到的重大困难，似乎又一次昭示了根本性理论问题的存在①。科学哲学，乃至哲学本身，在面临这样紧要的关头时，是否应该再次走出自身的领地，复兴自然哲学昔日的辉煌，并与理论科学和实验科学携手并进，为突破现存的科技创新“瓶颈”作出应有的贡献?

哲学与科学研究的相互交叉其实在当今国际舞台上已似星星之火，到处可见。无论是在欧洲还是在北美，为促进哲学与科学合作的机构，颇具影响力和国际声誉的就有好几所（如宇宙学哲学研究平台、哈佛大学的“黑洞创新”研究中心等）；而学者因为研究兴趣而自由形成的交叉研究团队，则数量更多。这种交叉式的、打破传统学科的做法，还不仅仅是出现在研究人员的研究工作当中，有些大学干脆就取消了本科生攻读的传统专业划分，让学生自己和指导老师一起制定能尽快衔接不断涌现的新发现、新发明的“交叉专业”。一时间，打破传统专业、重组教研方向的热潮此起彼伏。需要注意的是，这种打破传统专业领域的做法背后秉承的思想精神，是和那些为了市场需求、就业需要而进行的所谓“切实际、接地气”的学科专业改造思路完全不同的。

1 黑洞物理学与早期宇宙

由于2020年诺贝尔物理学奖颁发给了3位研究黑洞的科学家，黑洞和早期宇宙（或宇宙大爆炸理论）等相关话题再次引发公众关注。这次授奖显然是对这些领域在理论和实验方面长期以来积累的丰硕成果的承认。罗杰·彭罗斯（Roger Penrose）爵士获得了这次物理学奖金的一半。彭罗斯对国际哲学界的学者来说并不是陌生人，他本人常年关注哲学家对很多重要议题的探讨，这些议题包括广义相对论和量子力学之间的不相容问题、量子测量问题、意识的本质问题等；不仅如此，他还严肃认真地与哲学家们切磋合作，对解决这些哲学问题作出了重要贡献[6,7]。彭罗斯不仅是牛津大学哲学系的常客，也是横跨牛津、剑桥和纽约大学的“宇宙学哲学研究平台”（Research Platform on “Philosophy of Cosmology”）②的创立者之一。

黑洞和早期宇宙（或宇宙大爆炸理论）是一个哲学家特别感兴趣的科学话题。因为黑洞自从在概念上和数学上被推导出来后，它究竟是否存在，以及它的本质究竟是什么，始终备受争议。黑洞的概念最早来源于爱因斯坦的广义相对论场方程的奇异解。霍金和彭罗斯证明了这样的奇点不可能“赤裸”地存在于宇宙之中（彭罗斯给出的是更严格的证明），意思是

① Nima Arkani-Hamed，美国普林斯顿高等研究院的世界知名弦论学家，曾于2014年在加拿大圆周研究所（Perimeter Institute）的一次讲座中提到，高能物理学现在正在等待一场深刻的科学革命的发生。

② http://philosophy-of-cosmology.ox.ac.uk/.

说，它们总是被包裹“事件视界”（event horizon）之中——超过事件视界，任何物理事件都不再继续“发生”③，且任何能量都无法从事件视界内部逃脱[8,9]。然而，黑洞区域的内部和附近有巨大的引力效应，因此那里必然是广义相对论效应和量子效应同时发生显著作用之处。霍金提出的黑洞辐射就是一种发生在黑洞边界上的量子效应。可以说，如果未来人类真找到了一种把广义相对论和量子力学统一起来的精确理论（而不是近似理论），那么宇宙中没有哪里能比黑洞区域能更好地用来（尽管是间接地）检验该种理论的预测了。

彭罗斯[10]曾经提出过“将时空视为一种自旋网络”的扭量理论，该理论所开辟的进路后来演变成了圈量子引力论。在当今的高能物理学中，圈量子引力论是弦论的一个主要竞争者，两者都在更高的能量层级和更小的时空区域中超出了基本粒子物理学的标准模型。而圈量子引力的两位主要贡献者——Carlo Rovelli 和 Lee Smolin[11-13]，恰恰都有很高的哲学素养，也都是不少哲学论坛和会议的常客，这无疑是对反映于彭罗斯本人研究方法与学术成果中强烈的哲学倾向的一个印证。Rovelli 和 Smolin 在为非专业读者写的相关书籍中，对量子引力理论的那些令人困惑的特征给出了富有哲学深度的刻画，还提出了许多深刻的形而上学问题。不难看出，正因为近期内几乎不可能用实验判定量子引力的哪条理论进路是正确的，或仅仅是相对更好的，哲学在这里就不可避免地参与了进来。对于理论科学家来说，在选择理论假说时，传统力量和审美观起着重要的作用，这些因素从哲学观点看都与认知的价值（或方法的价值）紧密相连。其他一些关于黑洞和高能物理的挑战性问题则是形而上的问题，如：物理定律的失效问题，在黑洞的内部和在宇宙大爆炸之际（也许整个宇宙就是最大的黑洞？）的时空性质问题，物理学中的无穷大问题（比如在黑洞中心的物质密度是无穷大），以及关于真空中量子涨落的终极问题——最后这个问题引发了弦论学家关于宇宙的现状是来自于一种“微调”（fine-tuning）效应的猜测。这听起来似乎涉及了神灵的干预（换言之，可能是上帝亲手“微调”了整个宇宙，使之产生了宏观的物质和时空）。可以看出，这样的问题就是典型的哲学问题，因为它直接涉及宇宙学中的“人择原理”（anthropic principle）的正确与否——即，宇宙是不是必须如此这般地存在和演化，以至于使得宏观世界和人类能以目前被观测到的形式存在[14]？

前面提到的这个宇宙学哲学研究平台中，常职学者就有：数学家彭罗斯，物理学家 John Barrow、Joseph Silk，以及哲学家 Jeremy Butterfield、Simon Sounders、David Wallace。它主要由 John Templeton 基金会资助，是联结了世界顶级大学的高规格研究平台，也是汇集各界访问学者、承办众多学术会议和工作坊的具有国际领导地位的机构。在教学方面，它与牛津大学独有的本科专业“物理与哲学”④学位紧密相连。此外，国际上还有一个把黑洞物理研究与哲学相结合的重要中心——哈佛大学的“黑洞创新”（The Black Hole Initiative）研究中心⑤。这也是一个跨学科的研究中心，其成员包括天文学家 Sheperd Doeleman、数学家丘成桐、物理学家 Andrew Strominger 和哲学家 Peter Galison 等。这些机构的建设目标就是把科学家和哲学家汇集一堂，探索能够解决

③ 这是相对于远离黑洞区域的静止参照系而言；相对于黑洞内部系统的自身参照系而言，事件的发生并不会有什么特别异常的现象。

④ 这是一个提供实质性交叉的、独具特色的本科专业（而非一般性质的双学位）（https://www.ox.ac.uk/admissions/undergraduate/courses-listing/physics-and-philosophy#:~:text=The%20Oxford%20research%20group%20in,field%20theory%20and%20quantum%20gravity）。

⑤ https://bhi.fas.harvard.edu.

前面提到的那些重大问题的进路，以及探讨如何突破弦论或圈量子引力论在进一步发展中遇到的瓶颈。

其他着力于哲学与科学的交叉研究与教学的更大型的机构中，笔者认为还有 3 个非常值得关注：① 美国卡耐基梅隆大学的哲学系，包括 3 个分支——数学哲学与逻辑学、科学哲学与方法论、伦理学与价值论；其中，对医学伦理和人工智能伦理问题的研究在最后一个分支中表现得尤为出色⑥。② 德国慕尼黑大学的慕尼黑数理哲学中心（Munich Center for Mathematical Philosophy at LMU），旨在联合哲学与数学（更根本地说是指清晰、精密的方法论），来探究有关自然与实在的根本性问题⑦。③ 加拿大韦仕敦大学的罗特曼哲学研究所（Rotman Institute of Philosophy at Western University），倡导包括伦理与科学、物理学基础、心智与神经科学哲学、生态哲学等方面的研究⑧。

2 还原论、量子引力与时间问题

作为一种科学意识形态（scientific ideology），“还原统一论”有着很长的历史。常识告诉我们，世界中纷繁复杂的现象涉及许许多多在尺寸和尺度上迥异的系统。从星系、太阳系、恒星、行星、海洋、山峰、物种、躯体、器官、细胞、分子、原子直到基本粒子，不同尺度的系统对应着与其相应的科学分支，如天文学、天体物理学、地质学、动物学、生物学、化学和物理学等。这种意识形态背后的形而上学试图为它们提供这样的一种本体论，即认为一切系统都是由基本粒子构成，或者说是由量子场构成（由于物质有波粒二象性）；并且认为除了物理学之外的一切科学分支之所以有现在这样的样貌，归根到底是由于人类的感官知觉与心智能力的局限，而科学的一大目标就是统合所有的分支、并把它们还原为物理学（那种终极的、理想状态的物理学）[15]。这种意识形态最强烈的支持者就是逻辑实证主义，即前面提到的在 20 世纪初要把西方哲学从黑格尔哲学的统治中“拯救”出来的哲学运动。近些年来，这种“还原统一论”在科学哲学界已遭到批评和/或否定，承认各个科学分支的多样性和自主性已成为今日科学哲学界关于科学的一种新的意识形态[16]。然而，有讽刺意义的是，现在科学家们自己反倒开始逐渐打破学科边界，找到许多超越传统学科划分的共同的基础和方法。例如，如今的生物学家都经常要运用统计热力学的方法来解决生物学问题，而如果谁现在还对此表示出怀疑的态度，那他只会被认为是个外行。“还原”的主张也许依然不是个好的想法，但是“统一”的愿景已不再是一种空想了。

从哲学（或认知）的角度来看，物理学史上最深刻的成就之一就是麦克斯韦和玻尔兹曼建立统计热力学的工作。热力学是一种从对热现象的观察推广而得的科学：宏观变量（bulk variables）表征了宏观物质（bulk matter）在观察层面的状态，数学定律以静力学的风格（指平衡态过程）把这些变量联系起来。而统计力学则按照粒子的力学定律，概率地表征了大量微粒子组成的宏观系统，用微观细节严格说明了由概率决定的宏观状态的演变。这项工作之美妙就在于，处理同样的现象或观测结果，两种进路最终会完美地契合。知识论告诉我们，对于任何问题的研究都主要有两种方式：通过观察和归纳来研究和通过数学与推理来研究——前者是人类作为认知动物所必须依赖的、获取人类（或大型哺乳动物）生存尺度范围内为真的理论的方法，而后者才是有望带来超越人类自身尺度、在其他尺度范围内依然适用的关于自然物体与

⑥ https://www.cmu.edu/dietrich/philosophy/research/index.html.
⑦ https://www.mcmp.philosophie.uni-muenchen.de/about/index.html.
⑧ https://www.rotman.uwo.ca/activities/research-projects.

过程的理论的方法。统计热力学展现给我们的就是，这两种进路其实能以完美的相互协调和相互印证统一起来。所有其他把这两种进路结合起来或者把微观与宏观联姻起来的更精彩的科学事业，从哲学的角度来看，其实都是统计热力学的翻版[17]。

在科学前沿的种种挑战中，最新、最吸引人的，要算是量子引力理论了。量子引力的研究产生了各种各样的哲学困惑，从以下简单事实便可见一斑：直到20世纪早期，常识和物理学都告诉我们宇宙是由存在于时空并在时空中变化着的物质、能量、力所组成（也可以说是由各种物理场组成）；就像一部电影那样，物理过程历经时间，在电影的空间中发生着。但是后来，爱因斯坦创立了广义相对论，在那里，能量同作为背景的时空之间的区分被抛弃了——引力就是时空，时空就是引力；然而，为什么自然界4种基本相互作用中，只有引力是与时空等价的呢？更进一步说，其他3种基本相互作用都是量子化的，只有引力还不是，并且似乎无法被量子化。从形而上学的角度看，我们连量子化时空是什么意思都不完全清楚。能量量子化之后，能量子之间的跃迁在能量上是不连续的，但这种跃迁在时空中仍然是个连续的过程；如果连时空也被量子化了，时空量子（即引力子）之间的跃迁（即运动）究竟是什么又何以可能，就不仅仅是一个科学理论问题，更是一个为一般的科学与哲学常识所无法理喻的形而上学问题。因此，不难想象，作为时空的引力如果真能被量子化，就会产生出一系列严重的形而上学问题。

量子引力论中最广为人知的问题就是时间问题。就像Rovelli[18]提醒我们的，时间这种最通常的现象、这个我们一切的经历和记忆所处的背景，随着物理学研究的深入，已经逐渐“消失”了。当相对运动速度与光速可以相比拟时，两个参考系中理想钟表测出的时间就会以不同的速率流逝。“现在”在客观上也变成了一个依赖于参考系的东西。而当引力场足够强时，两个邻近的区域就会有不同的时间。另外，在考察自然的基本规律时会发现，除了一个特例（热力学第二定律）之外，其他所有基本自然律在时间反演变换下都是不变的，所以在自然的基础层面，其实没有过去和未来之分。我们所熟知的时间，即那种从未来到现在、再由现在到过去的一维、连续、统一的事件流，似乎只是某种东西的近似，而这种东西既没有确定的现在，也没有未来和过去之间的根本区别；甚至，在引力场足够强的情况下，都不具有一个统一的时序。此外，当时间作为量子引力场的一个方面被量子化之后，它甚至会变成不连续的。在沿着这条路线对时间的研究中，最令人吃惊的可能是下面这个事实：时间，乃至因果过程，均来源于物理学中唯一的一条不遵循时间反演对称性的定律——热力学第二定律，即孤立系统的熵只会增加而不会减少[19]。因果性和时间的方向都仅仅来自这条定律，而这条定律却并不属于自然界的基本定律；所有基本定律既不遵循时间的方向，也不遵循因果性⑨[20]。有讽刺意味的是，对我们人类来说，可能没有什么比时间和因果性更加可畏和“残暴”的了（而“命运”，不正是二者的结合吗？），但这两者都只不过是大自然在人类所处的尺度上的统计性事件而已。还有什么比这样一个事实更能告诉我们，人类在大自然中是多么的脆弱？

事实上，量子引力论的场方程中时间变量已经不复存在。对科学其他任何分支的研究者来说，时间变量是理论或方程中最常见到的变量，是用来表述运动

⑨ 物理学中的基本定律方程都是既可以从因推导出果，也可以从果推导出因，区别不过是两组条件的相互可导出性。任何这样两组条件中，孰是因、孰是果，不是用基本定律方程可以决断的——这就是为什么大哲学家、逻辑学家罗素在他的著名论文《关于原因的概念》中强调在基础科学领域中因果性是不存在的。

和变化的最基本的参量，但量子引力方程中却没有它。原因不难理解：如果时间被量子化，那么就存在时间的“量子”（作为引力子的一部分）；而根据相对论，时间量子之间又必须是相对的，那么它们就都必须是定域的（local），并具有量子涨落性；因此，那个经典的、连续的时间变量（把连续的时刻作为其取值的变量）当然就在量子引力方程中变得没有意义、没有地位。与此同时，广义相对论是把宏观的时空作为研究对象的，而且我们的确感受到了连续的时间，以及跨越空间的时间切片（time slices）序列。是否由此可以猜测，量子引力论中的时空量子和广义相对论中连续时空之间的关系，正是统计力学和热力学之间的关系[21]？

3 知觉、智能与自由能原理

本文第 2 节说到的统一理论[22]和本节要谈的统一理论相比，可谓相形见绌。如果 Friston 等[23]提出的自由能（最小化）原理（Free Energy [Minimization] Principle，FEP）最终被证明为真或近似为真，那么动物的知觉、行动、智能等现象都可以在“自由能最小化”这条自然原理中得到解释，甚至生物的演化过程也可以由该原理解释。自由能原理有希望在整个自然科学中扮演牛顿力学在 20 世纪之前的物理学中扮演过的角色。由于这方面的研究与突破最早是在动物的知觉智能（perceptive intelligence）研究中实现的，而知觉智能又是人工智能与深度机器学习研究所关注与模拟的领域，因此动物知觉智能的预测处理（predictive processing）模型，即所谓的“PP 模型”或“预测心灵论”，可谓是当今研究智能有关的领域内最热的话/课题了。

有关知觉的 PP 模型和生命系统的 FEP 的科研工作已经引起了哲学家们的关注，其中 Clark[24]和 Hohwy[25]是成就最突出的两位。Arbib[26]则是一位更偏重科学方面，但哲学上也毫不逊色的多才多艺的科学家，其研究涉及计算机科学、认知科学、神经科学和相关的哲学；他甚至还跟在科学模型方面做出先驱性工作的科学哲学家 Hesse 合作出版过一本名为《实在的建构》[27]的专著——这个书名正好与 PP 模型关于人脑如何感知外部环境的说法吻合。

依据 Friston 等关于 FEP 的工作，Clark 和 Hohwy 拓展了 PP 模型的理论和实验成果，用它覆盖了整个的动物知觉与认知领域。PP 模型论把动物知觉与行动看作是一个整体性的、基于贝叶斯式的预测误差最小化（Bayesian prediction error minimization）的系统⑩，以此拒斥了传统的“感觉—思考—行动”序列模型，从而导向了认为生命系统具有多面性和整全性的观点。它同时还打开了解决各类“起源问题”的可能性。比如，关于智能、认知和语言是如何通过进化而演生出来的问题。动物学习也不再是一个孤立被动的、总是发生在决策和行动之前的过程；相反，学习、审度、决策和行动，都被看作是跟主体所在的那个环境进行互动的不同形式。不难看出，这个整全模型提供了一个更好的框架，可以用来考察智能或语言是如何经由自然选择而逐渐演化、衍生出来的，以及考察如像眼睛这样的原始感觉器官是怎么从上述功能的整合发展中演化出来的。智能的演生及其他的高阶动物官能的演生是生命科学的一个前沿领域，吸引着科学家，也吸引着哲学家。

尽管这种“统一”如果能被实现将是颇为令人激动之事，但它真的能实现吗？动物知觉从根本上讲，真的是一种主动推理吗？当以 PP 模型来理解时，知觉真的就是行动、行动真的就是知觉吗？简单地说，PP

⑩ PP 模型下的动物知觉系统是一个非意识的但遵循着贝叶斯概率更新方案（updating scheme）的系统；该方案的工作过程是，首先产生对环境的预测，然后控制并最小化由预测与感官输入的冲突所产生的误差项（error terms），最后再将无法消除的错误送回上层大脑皮层的过程。PP 模型的感知预测误差最小化原理是理论上更抽象、更一般的自由能最小化原理在动物感知系统上的应用。

模型下的动物知觉不再是一个“加工厂”式的系统，把从感官摄入的信息“加工”成知觉判断，而是一种像“贝叶斯机器”（Bayesian Machine）式的东西：在其运行中，大脑对其环境做出预测、同时让肢体干预该环境（即行动），两者按照大脑自由能的 FEP 协同作用，使大脑最后形成真实反映其环境的知觉判断。另外，镜像神经元的存在真是 FEP 的结果吗？也就是说，在哺乳动物大脑里出现的具有行动和观察的镜像效应（mirror effect of action and observation）的神经元，是从感知与行动为了使大脑的自由能最小化而做出的整合行为中演生出来的吗⑪？这里虽然有一些属于经验问题，需等待科学探究来验证或否证，但其他的显然属于哲学问题。知觉的 PP 模型的核心——“主动推理”或“贝叶斯式推理”中的“推理”概念到底是什么？哲学中有一个为人熟知的模型——波普尔的科学假设验证模型，它基于一种由提出假设到检验假设的逻辑推理模式。当这一波普尔模型配备了贝叶斯推理规则，它就变得跟 PP 模型的知觉原则极为相似；但是波普尔模型毕竟是一种科学方法，而 PP 模型则是一种动物大脑中无需意识或认知的、可以自动运行的知觉过程（不需利用概念和推理来构造假设、检验假设，然后决定接纳或拒绝它们）。要说麦克斯韦方程组描述了电磁场的行为，这应该没有什么问题；但是要说一个像贝叶斯化的波普尔科学方法论模型描述了几乎仅涉及肉体运动的大脑中的神经突触和神经元的行为，这又是什么意思呢？当然，也许这样大的转换和跨度并非不可能，只不过是哲学家由于自身缺乏想象力才拒绝接受 PP 模型的形而上学可靠性，也就是拒绝想象能量在大脑中的无意识转移过程为什么会遵循一些反映了科学方法的规则。正如本文第 4 节中要谈的，如果昆虫甚至癌细胞都可以“采取策略”或者“博弈”以在生存竞争中获胜，那么为何大脑不能用“推理”来“猜测”它遇到的有可能产生感官输入的外部原因呢？

4 演化论与博弈论

“演化博弈论”（evolutionary game theory）是一个通过结合演化生物学与博弈论而产生的发展迅速的交叉学科。Maynard-Smith[28]是 20 世纪 50 年代最杰出、最知名的演化博弈论创始人。他把经济学中的博弈概念引入生物学，建立了从生物单位（如细胞、昆虫、人类等）之间的博弈压力与选择来研究生物表现型演化的研究分支。接下来的“里程碑”大概要数 Sigmund[29]和 Nowak[30]的工作了。Nowak 统一了博弈演化论的理论，并把其应用扩展到空间博弈（spatial games）和其他具有复杂结构的群体⑫。

人类是社群性的动物，个体的存在和生活在大部分情况下依赖于跟他人的关系。尽管“游戏/博弈”⑬这个术语一般指的是为了娱乐或赢钱的玩家策略，但科学家起初用它意指经济和政治中有策略的交锋，后来又意指一切需要用策略来智胜对方、获得收益的关系。博弈论就是研究这种关系的数学理论。如 Maynard-Smith 在演化博弈论中发现的，如此理解的博弈也可以在动物群体甚至益生元群体中得以应用。在博弈论中，普遍假定博弈各方具有为获取利益（utilities）做出理性选择的能力；而在演化博弈论

⑪ “镜像神经元”假设是指，在动物大脑中发现的一组神经元，无论是在动物自己执行某个动作时，还是在看到另一只动物执行同样的动作时，该组神经元都会被激活。这就意味着动物的行为和观察二者之间有着紧密的联系；从心理的角度，该发现提供给我们理解“他心”问题的线索，而从认知的角度，它似乎说明动物在认知其环境时，其大脑皮层里进行的自由能最小化过程（即上面说到的 PP 模型中的预测误差最小化）是通过观察和行动的相互结合来实现的。比如，在观察事物时，我们移动眼睛、肢体、使用工具等，这些都是通过预测与行动二者的协同来得到自由能或预测误差的最小化。

⑫ 演化博弈论最为典型的应用是在大型、均匀、随机的生物群体中。引入群体中个体的非均匀空间分布，会给问题和方程的解析推演带来各种各样的困难。

⑬ 英文中“游戏”和“博弈”均为“game”。

中，仅要求个体具有作为自然选择机制的差异适应度（differential fitness）。人与人之间的博弈由各人理性地根据获利（payoff）和概率来运算、选择策略，然后以获利值的高下来决定胜负；而在生物体与生物体之间，则是个体在博弈竞争中运用了使得个体在自然环境中繁衍出更多下一代的策略而取胜，这里所谓的策略并不需要有任何理性选择的含义。此外，当具有理性的主体展开重复博弈的过程时，自然演化过程中的物种间博弈也会在人类社会中发生。无论是否理性、是否采用决策，这种重复的博弈过程经过一段时间，总会得到一个像是经历了自然选择而得出的结果。演化与决策在博弈中相互对比着出现，前者是自然的，后者是理性的。当博弈论从经济学转移到生物学时，演化取消了理性，得到了演化博弈论；而当演化博弈论重新回到经济学和政治学中时，演化的过程又取消了决策者的理性。在 Nowak 的演化动力学时代，这个理论覆盖了所有种类和尺度的生物种群，包括人类社会。生命的核心方程就是演化博弈论中的复制体方程（replicator equations）和数学上与之等价的种群动力学中的洛特卡-沃尔泰拉方程（Lotka-Volterra equations）。这一数学理论已经逐步得到展开，并用于更广泛的现象[31]。

伦敦政治经济学院的哲学教授、哲学系主任 Alexander 多年来开展的研究工作就是从演化博弈论的角度考察道德的演化。Alexander[32]所著的《道德的结构演化》是该领域中影响甚大的一本著作。利用演化博弈论的论证和洞见，他为元伦理学的诸多问题提供了新的视角和解决方案。在西方哲学传统中，关于人们对公平、正义的感受和辩护，以及对正确与错误的道德感的本质，霍布斯和康德被认为是站在立场对立的两端。霍布斯式的道德理论是建立在“个体的根本自私性和野蛮性”与群居条件下“分享有限资源的必要性”这两个相悖的事实上的；而康德式的道德理论则诉诸存在于每个个体的理性灵魂中的理性原则或律令（imperatives）。无论对于二者中的哪一个（甚至对于任何其他讨论道德的传统进路），Alexander 的演化博弈论思路都可谓是一种改进；其关键就在于，该思路悬置了一切关于动机（我们是不是本质上自私）和理性（我们做出的决策是优还是劣）的事项，而让社会条件去约束这些根据所选的“策略套餐”（strategy set）而反复进行的博弈过程，看它按照复制者方程[30]如何进行演化。人们已经在计算机中做了这类社会博弈的模拟。获胜策略最终将占据全部群体（或几乎全部群体）。结果表明，公平、正义等，恰好是在最大的一组可相互替代的相容条件下，大多数跟人类群体相似的群体所发展的策略。例如，在“分蛋糕”的博弈模拟中，不均等分割策略虽然可能会占据全部群体，但是这种情况下所需的对群体的约束，比起得到均等（也就是公平）分割策略的约束来说，是严之又严。我们难道不能猜测，这就是作为道德准则的“公平”从人类社会的长期演化中产生出来的过程？

关于道德的演化的研究工作现在已经和关于知觉与智能的研究工作结合起来（参见第 2 节），这种交叉研究的结果之一就是得到了对“天生观念”（innate ideas）的存在和本质的洞察，而“天生观念”曾在西方哲学史中饱受争议。人类是不是生下来就有一些先验为真的观念？哲学史上有两个比较出名的回答：笛卡尔认为“是”，而洛克认为“否”。但是，如果上述研究揭示了我们关于公平和正义的感受或认识是生物的和文化的演化产物，那么这种感受可能就是对公平和正义的天生感受或认识。这种思路如果进一步走下去就可以问，我们对于概率的认识、对于“诸多中立选项应该具有相同概率”的认识，是先天的还是学习来的？已经有可靠的婴儿心理学实验指出了这些观念的天生特征[33]。同样的，根据 PP 模型，人类的知觉用以构建关于所处环境的知觉判断的贝叶斯规则是先天的（*a priori*）还是天生的（innate）？或者可以问，

对先天/后天（*a priori/a posteriori*）、对分析/综合的传统划分，是不是已经过时？包括这些在内的很多问题，都是今天的科学家和哲学家在学科交叉共同体中提出来并正在进行严密研究的。

如果有人认为演化博弈论并不考虑竞争个体是理性的人还是昆虫，因而把它的方法拿来用于人类社会显得太激进了，那么认为自然界中不仅人类这个物种有主体性（agency），而且“主体性无处不在”可能就更显激进了。目的论解释（teleological explanations）曾经是自然哲学中一类合理的解释，在亚里士多德哲学传统中尤其是这样。自牛顿力学成为西方主流世界观之后，这类解释就显得可疑了。的确，演化类型的解释进路，或者说用自然选择来解释演化的进路，跟牛顿式的、爱因斯坦式的、量子力学式的解释进路都不相同；但即使是在这里，目的（*telos*）或主体性（即为了理由或目的而做出的行动）似乎也不应该有地位，尽管大自然对物种的选择似乎是为了某种理由或目的。然而，近来情况有了变化，有赖于对“决策”和“行动”的跨领域交叉科学思考，科学家已经突破了这个长期在哲学家眼中神圣不可触动的边界。

根据 Okasha[34]的说法，生物学和社会科学中大量关于演化过程的证据表明，主体性和目的性时常被用于构建模型和提供解释。在生物学研究的实践中，根据他的说法，主体性的概念出现了两种形式：一种适用于生物体个体，另一种适用于自然界或演化和适应的环境。生物物种究竟是会演化到一个确定的终点，还是（用有价值意味的话来说）由于具备了某种最优特征才存活下来，现在仍是个有争议的话题。这种争议就像是对于为生存而作对的两种甲虫，我们能不能说“选择较大的身躯还是选择较硬的甲壳”是它们用以博弈的策略呢？另一个相关的概念源于“人类的理性可以从利益最大化的角度来理解”这种观念。一个理性的行动就是在其他条件均同时给行动者带来最大利益的行动。博弈中的行动主体就是用尽一切手段追求最佳结果的主体。鉴于演化和最优化之间可能存在的这种对称关系，我们也许可以论证，理性作为人类的基本特征之一，正是长期演化的产物。同样是Okasha 指出的，这种联系可以从两种角度来看：一种角度是认为理性本身就是适应环境（adaptation）的产物，另一种角度是“为理解其行为，我们完全可以把历经了演化的生物体看成是有理性的”。当然，究竟哪种是正确的，依赖于“适配最大化”（fitness maximization）能不能和“效用最大化”（utility maximization）相符——前者是关于适应的概念，从而是关于自然选择的概念；而后者是关于理性的概念，从而也是关于主体能力和行为准则的概念。

5 总结

本文选择性地报告了笔者认为非常重要、同时也略微知晓的例子，希望能抛砖引玉，激发广大科学界和哲学界的读者对当下一些值得关注且很有意思的话题的兴趣。笔者试图指出，基础科学中极富思维挑战的根本性概念难题需要哲学、形而上学的洞见，生命科学与社会科学的互动研究中也需要认识论、伦理学的参与；与此同时，所有这些蓬勃发展的科学探索也给传统哲学难题的讨论带来了诸多意想不到的启迪和推进。在目前这个科学又将迎来大变革的时代，科学与哲学的相互启发和深度联盟，有助于我们深入探索各学科中重大的基础性难题，有助于我们发展出拥有跨学科特点的新的研究方法，从而或许有助于我们将来创造出具有全新境界的科学。

特别需要声明的是，本文完全没有提及关于类脑智能与人工智能的进展，不是因为该领域哲学与科学鲜有交叉，而是因为这里的交叉研究太多、太醒目，笔者在这里很难以合适的篇幅来讨论，只好选择完全回避的策略。

参考文献

1 Newton I, Hall A R, Tilling L. The Correspondence of Isaac Newton. Volume 7. Cambridge: Cambridge University Press, 2020.

2 Westfall R. Never at Rest: A Biography of Isaac Newton. Cambridge: Cambridge University Press, 1980.

3 Pais A. Subtle is the Lord (edition with Penrose's forward). Oxford: Oxford University Press, 2005.

4 Taylor C. Hegel. Cambridge: Cambridge University Press, 1977.

5 Richardson A, Uebel T. The Cambridge Companion to Logical Empiricism. Cambridge: Cambridge University Press, 2007.

6 Penrose R. Shadows of the Mind. Oxford: Oxford University Press, 1996.

7 Penrose R. Emperor's New Mind. Oxford: Oxford University Press, 2002.

8 Penrose R. Gravitational collapse and space-time singularities. Physical Review Letters, 1965, 14(3): 57.

9 Hawking S W. Black hole explosions? Nature, 1974, 248: 30-31.

10 Penrose R. Applications of negative dimensional tensors// Combinatorial Mathematics and its Applications. London: Academic Press, 1971.

11 Rovelli C, Smolin L. Loop space representation of quantum general relativity. Nuclear Physics B, 1990, 331(11): 80-152.

12 Rovelli C. Reality is Not What It Seems: The Journey to Quantum Gravity. New York: Riverhead Books, 2018.

13 Smolin L. Three Roads to Quantum Gravity. New York: Basic Books, 2008.

14 Donoghue J F. The fine-tuning problems of particle physics and anthropic mechanisms// Carr B, ed. Universe or Multiverse. Cambridge: Cambridge University Press, 2007: 231-246.

15 Oppenheim P, Putnam H. Unity of science as a working hypothesis// Boyd R, Gasper P, Trout J D, eds. The Philosophy of Science. Cambridge: The MIT Press, 1991: 403-427.

16 Cartwright N. The Dappled World: A Study of the Boundaries of Science. Cambridge: Cambridge University Press, 1999.

17 Jaynes E T. Probability Theory: The Logic of Science (The posthumous work of Jaynes). Cambridge: Cambridge University Press, 2003.

18 Rovelli C. The Order of Time. New York: Riverhead Books, 2018.

19 Callender C. What Makes Time Special? Oxford: Oxford University Press, 2017.

20 Russell B. I.—On the notion of cause. Proceedings of the Aristotelian Society, 1913, 13(1): 1-26.

21 Rovelli C. The Disappearance of Space and Time. Philosophy and Foundations of Physics, 2006, 1: 25-36.

22 Barrow J. New Theory of Everything. Oxford: Oxford University Press, 2007.

23 Vossel S, Geng J J, Friston K J. Attention, Predictions and Expectations, and Their Violation: Attention Control in the Human Brain. Frontiers in Human Neuroscience, 2014, 8: 490.

24 Clark A. Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford: Oxford University Press, 2016.

25 Hohwy J. The Predictive Mind. Oxford: Oxford University Press, 2014.

26 Arbib M. How the Brain Got Language. Oxford: Oxford University Press, 2012.

27 Arbib M, Hesse M. The Construction of Reality. Cambridge: Cambridge University Press, 1986.

28 Mayard-Smith J. Evolution and the Theory of Games. Cambridge: Cambridge University Press, 1982.

29 Hofbauer J, Sigmund K. Evolutionary Games and Replicator Dynamics, Cambridge: Cambridge University Press, 1998.

30 Nowak M. Evolutionary Dynamics: Exploring the Equations of Life. Harvard: The Belknap Press of Harvard University Press,

2006.

31 Hofbauer J, Sigmund K. Evolutionary Games and Replicator Dynamics. Cambridge: Cambridge University Press, 1998.

32 Alexander J M. The Structural Evolution of Morality. Cambridge: Cambridge University Press, 2007.

33 Strevens M. Tychomancy: Inferring Probability from Causal Structure. Harvard: Harvard University Press, 2013.

34 Okasha S. Agents and Goals in Evolution. Oxford: Oxford University Press, 2018.

Review of Recent Developments of International Interdisciplinary Research in Philosophy and Science

LIU Chuang[1*] ZHU Kefu[2]

(1 School of Philosophy, Fudan University, Shanghai 200433, China;

2 College of Humanities, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract In this article we discuss a selected sample of interdisciplinary researches between philosophy and science in the forefront of high energy physics, science of intelligence, biological science, and social science. We discuss both the ideas and the important international institutions and figures.

Keywords Interdisciplinary, blackhole physics, early universe physics, quantum gravity, predictive processing model, free energy principle, evolutionary game theory



刘 闯 复旦大学哲学学院科学哲学与逻辑学系教授、智能科学与智能哲学中心主任，特聘教授、博士生导师。美国匹兹堡大学科学史与科学哲学系博士。曾任美国佛罗里达大学哲学系教授。主要从事科学哲学、物理哲学、智能哲学等方面的研究工作。在国际科学哲学顶级刊物上发表多篇论文。曾获美国国家自然科学基金（NSF）项目与社会科学基金（NEH）项目，获中国教育部高等学校科学研究优秀成果奖（人文社会科学）一等奖。E-mail: chuangliu@fudan.edu.cn

LIU Chuang Distinguished Professorship at Fudan University; Doctoral Supervisor at the Department of Philosophy of Science and Logic, School of Philosophy; Director of the Center for the Science and Philosophy of Intelligence. He was a full professor in the Department of Philosophy at the University of Florida since 2005, and he received his Ph.D. from the Department of HPS at the University of Pittsburgh. He specializes in philosophy of science, philosophy of physics, and philosophy of intelligence, has published many articles in the top international journals in philosophy of science. He has leaded projects funded by National Science Foundation and by the National Endowment for the Humanities of the United States, and has received the First Prize of Scientific Research Achievement Award (Social Science Division) of the People’s Republic of China of Education Ministry in 2020. E-mail: chuangliu@fudan.edu.cn

■责任编辑：岳凌生

*Corresponding author